A catalog of genetic loci associated with kidney function from analyses of a million individuals

LifeLines Cohort Study; Million Veteran Program; Wuttke, Matthias; Li, Yong; Li, Man; Sieber, Karsten B.

Published in:
Nature Genetics

DOI:
10.1038/s41588-019-0407-x

Publication date:
2019

Document Version
Peer reviewed version

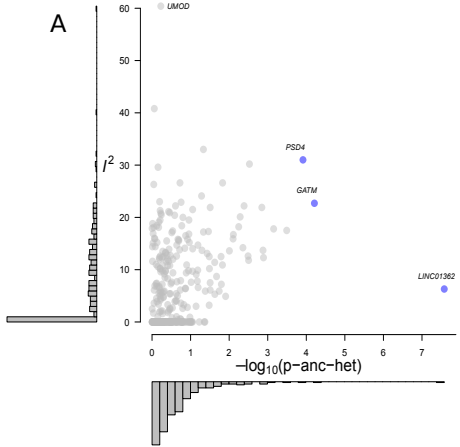
Figure 1 – Trans-ethnic GWAS meta-analysis identifies 308 loci associated with eGFR

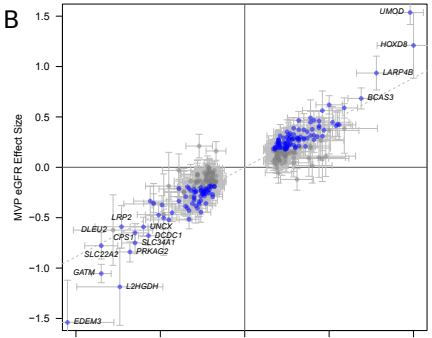
Circos plot: Red band: $-\log_{10}(P)$ for association with eGFR, by chromosomal position. Blue line indicates genome-wide significance ($P=5\times10^{-8}$). Black gene labels indicate novel loci, blue labels known loci. Green band: Measures of heterogeneity related to the index SNPs associated with eGFR. Dot sizes are proportional to I^2 or ancestry-related heterogeneity (p-anc-het). Blue band: $-\log_{10}(P)$ for association with CKD, by chromosomal position. Red line indicates genome-wide significance ($P=5\times10^{-8}$). Radial lines mark regions with p-anc-het < 10^{-3} or I^2 >25%. Inset: Effects of all 308 index SNPs on log(eGFR) by their minor allele frequency, color-coded by the associated odds ratio (OR) of CKD (red scale for OR≤1, blue scale for OR>1). Triangles highlight SNPs that were significantly ($P<1.6\times10^{-4}$ =0.05/308) associated with CKD.

eGFR
I²
p-anc
CKD
log(eGFR) Effect Size
Minor Allele Frequency
CKD OR/p−value
0.80–0.85
0.86–0.90
0.91–0.95
0.96–1.00
1.01–1.05
1.06–1.10
1.11–1.15
1.16–1.20
p < 1.6 · 10⁻⁴
p ≥ 1.6 · 10⁻⁴

Figure 2 – Generalizability with respect to other populations and other kidney function markers

Panel A: Measures of heterogeneity for 308 eGFR-associated index SNPs. Comparison of each variant's heterogeneity quantified as I^2 from the trans-ethnic meta-analysis (Y-axis) vs. ancestry-related heterogeneity from meta-regression ($-\log_{10}$(p-anc-het), X-axis). Histograms summarize the distribution of the heterogeneity measures on both axes. SNPs with significant p-anc-het ($<1.6\times10^{-4}$ = 0.05/308) are marked in blue and labeled, SNPs with I^2 >50% are labeled. **Panel B: Comparison of genetic effect sizes between CKDGen Consortium data (X-axis) and MVP data (Y-axis).** Blue font indicates *P* $<1.6\times10^{-4}$ (0.05/308) in the MVP. Error bars indicate 95% CIs. Dashed line: line of best fit. Pearson's correlation coefficient: 0.92 (95% CI: 0.90; 0.94). **Panel C: Comparison of the magnitude of the effects on eGFR (X-axis) vs. BUN (Y-axis) for the 308 eGFR-associated index SNPs.** SNPs are marked in blue when $P<1.6\times10^{-4}$ (0.05/308) in the BUN analysis. Error bars indicate 95% CIs. Dashed line: line of best fit. Pearson's correlation coefficient: -0.66 (95% CI: -0.72; -0.59).


A
UMOD
PSD4
GATM
LINC01362
I^2
$-\log_{10}$(p-anc-het)


B
UMOD
HOXD8
LARP4B
BCAS3
LRP2
DLEU2
UNCX
CPS1
DCDC1
SLC34A1
SLC22A2
PRKAG2
GATM
L2HGDH
EDEM3
MVP eGFR Effect Size
CKDGen log(eGFR) Effect Size

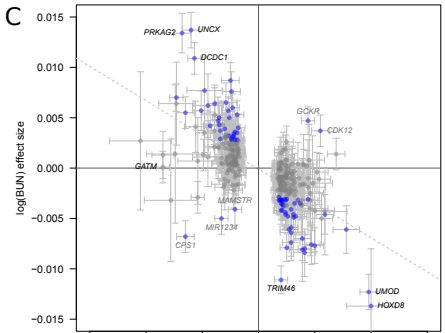

C
PRKAG2
UNCX
DCDC1
GCKR
CDK12
GATM
MAMSTR
MIR123
CPS1
TRIM46
UMOD
HOXD8
log(BUN) effect size
log(eGFR) effect size

Figure 3 – Human orthologs of genes with renal phenotypes in genetically manipulated mice are enriched for association signals with eGFR

Signals in candidate genes identified based on the murine phenotypes abnormal GFR (**Panel A**), abnormal kidney physiology (**Panel B**), and abnormal kidney morphology (**Panel C**). Y-axis: $-\log_{10}(P)$ for association with eGFR in the trans-ethnic meta-analysis for the variant with the lowest p-value in each candidate gene. Dashed line indicates genome-wide significance ($P=5\times10^{-8}$), solid gray line indicates the significance threshold for each nested candidate gene analysis (included in lower right corner in each panel, experiment-wide significance). Orange color indicates genome-wide significance, red color experiment-wide but not genome-wide significance, and blue color indicates genes with no significantly associated SNPs. Genes are labeled when reaching experiment- but not genome-wide significance; black font for genes not mapping into loci reported in the main analysis, gray font otherwise. Enrichment p-value reported for observed number of genes with association signals below the experiment-wide threshold vs. the expected number based on the complementary cumulative binomial distribution (Methods).

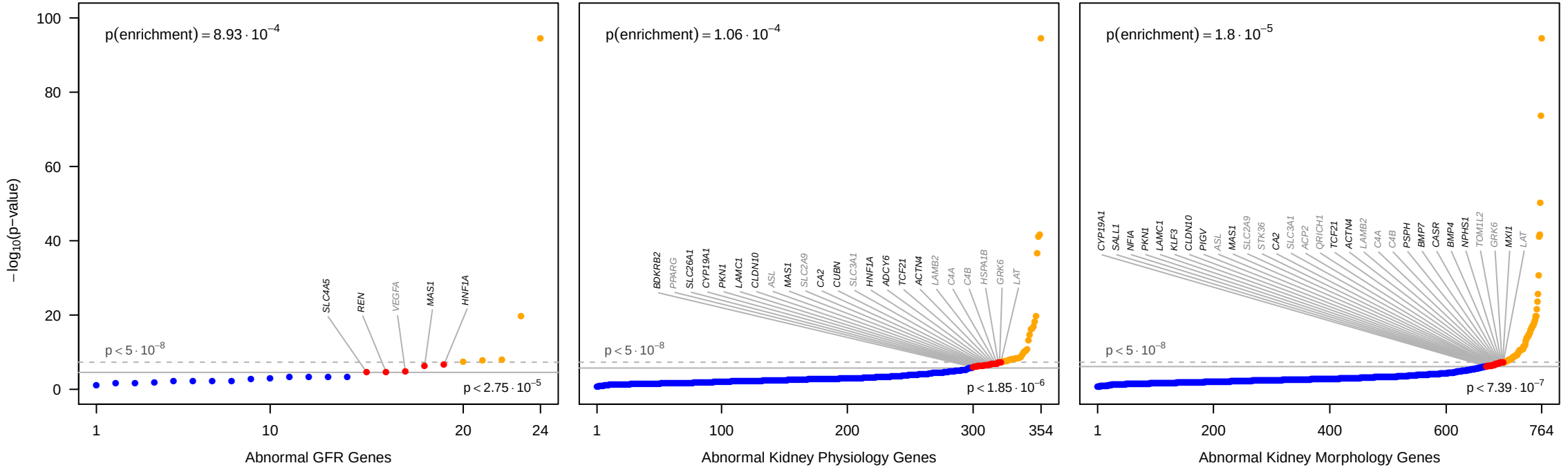

p(enrichment) = 8.93 · 10^−4
p < 5 · 10^−8
p < 2.75 · 10^−5
SLC4A5
REN
VEGFA
MAS1
HNF1A
Abnormal GFR Genes
p(enrichment) = 1.06 · 10^−4
p < 5 · 10^−8
p < 1.85 · 10^−6
BDKRB2
PPARG
SLC26A1
CYP19A1
PKN1
LAMC1
CLDN10
ASL
MAS1
SLC2A9
CA2
CUBN
SLC3A1
HNF1A
ADCY6
TCF21
ACTN4
LAMB2
C4A
C4B
HSPA1B
GRK6
LAT
Abnormal Kidney Physiology Genes
p(enrichment) = 1.8 · 10^−5
p < 5 · 10^−8
p < 7.39 · 10^−7
CYP19A1
SALL1
NFIA
PKN1
LAMC1
KLF3
CLDN10
PIGV
ASL
MAS1
SLC2A9
STK36
CA2
SLC3A1
ACP2
QRICH1
TCF21
ACTN4
LAMB2
C4A
C4B
PSPH
BMP7
CASR
BMP4
NPHS1
TOM1L2
GRK6
MXI1
LAT
Abnormal Kidney Morphology Genes
−log10(p−value)
100
80
60
40
20
0
1
10
20
24
1
100
200
300
354
1
200
400
600
764

Figure 4 – Credible set size (X-axis) vs. variant posterior probability (Y-axis) of 4,060 variants in 212 99% credible sets by annotation

Panel A: Exonic variants. Variants are marked by triangles, with size proportional to their CADD score. Red triangles and variant labeling indicate missense variants mapping into small (≤5 SNPs) credible sets or with high individual posterior probability of driving the association signal (>0.5). **Panel B: Regulatory potential.** Symbol colors identify variants with regulatory potential as derived from DNAse hypersensitivity analysis in target tissues (Methods). Variant annotation was restricted to variants with variant posterior probability >1%; SNPs with posterior probability ≥90% contained in credible sets with ≤10 variants were labeled.

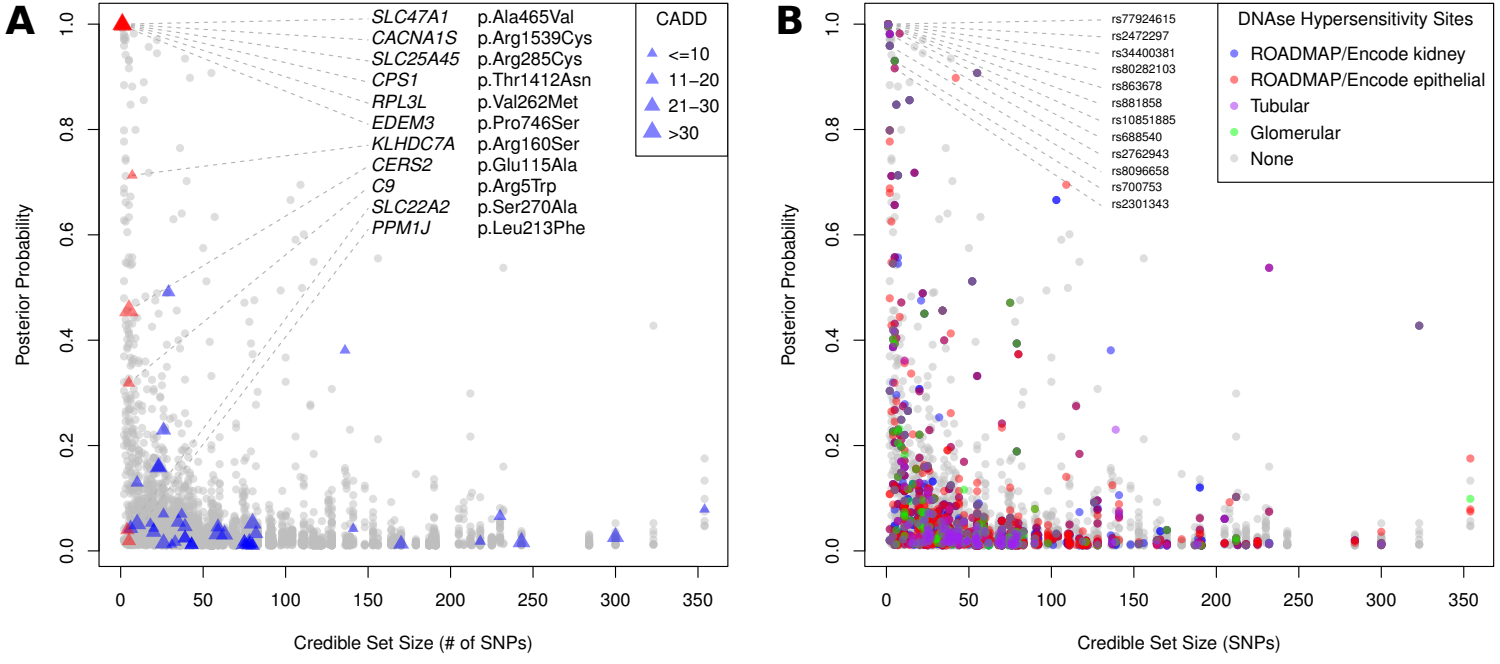

A
SLC47A1 p.Ala465Val
CACNA1S p.Arg1539Cys
SLC25A45 p.Arg285Cys
CPS1 p.Thr1412Asn
RPL3L p.Val262Met
EDEM3 p.Pro746Ser
KLHDC7A p.Arg160Ser
CERS2 p.Glu115Ala
C9 p.Arg5Trp
SLC22A2 p.Ser270Ala
PPM1J p.Leu213Phe
CADD
<=10
11–20
21–30
>30
Posterior Probability
Credible Set Size (# of SNPs)
B
rs77924615
rs2472297
rs34400381
rs80282103
rs863678
rs881858
rs10851885
rs688540
rs2762943
rs8096658
rs700753
rs2301343
DNAse Hypersensitivity Sites
ROADMAP/Encode kidney
ROADMAP/Encode epithelial
Tubular
Glomerular
None
Posterior Probability
Credible Set Size (SNPs)

Figure 5 – Co-localization of eGFR-association signals with gene expression in kidney tissues

All eGFR loci were tested for co-localization with all eQTLs where the eQTL cis-window overlapped (±100 kb) the sentinel genetic variants. Genes with ≥1 positive co-localization (posterior probability of one common causal variant, H4, ≥0.80) in a kidney tissue are illustrated with the respective sentinel variants (Y-axis). Co-localizations across all tissues (X-axis) are illustrated as dots, where the size of the dots indicates the posterior probability of the co-localization. Negative co-localizations (posterior probability of H4 <0.80) are marked in grey, while the positive co-localizations are color-coded based on the predicted change in expression relative to the allele associated with lower eGFR.

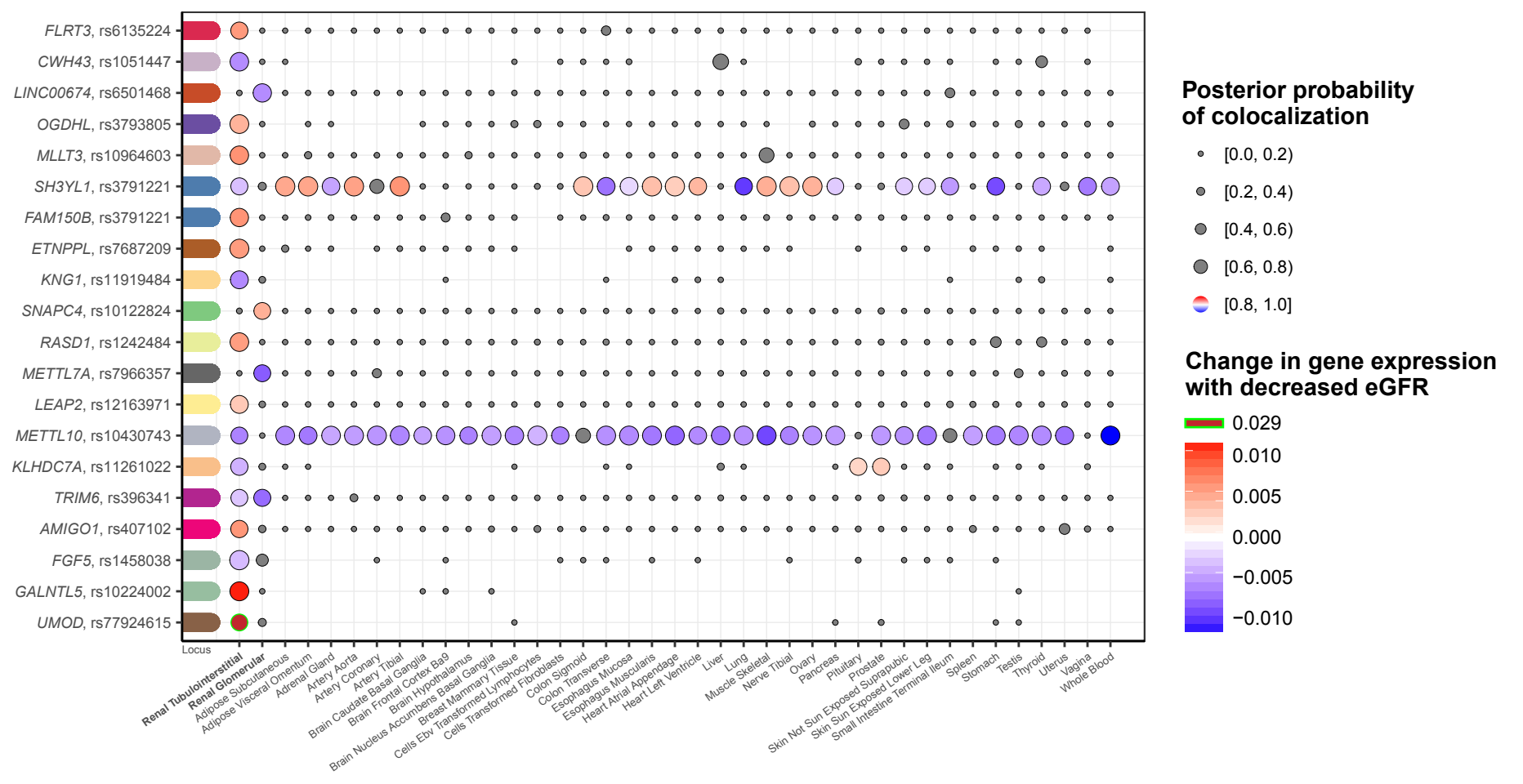
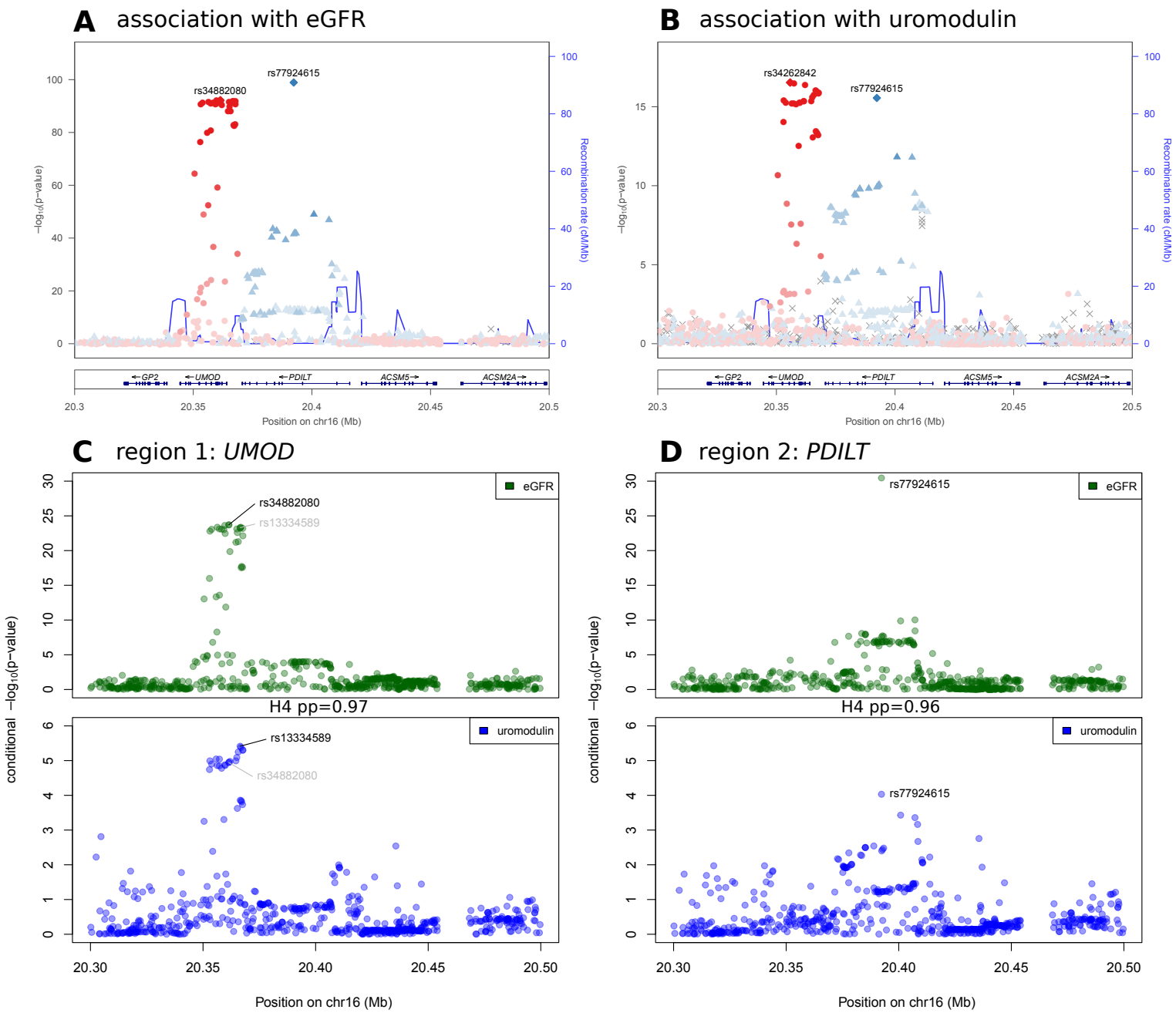

FLRT3, rs6135224
CWH43, rs1051447
LINC00674, rs6501468
OGDHL, rs3793805
MLLT3, rs10964603
SH3YL1, rs3791221
FAM150B, rs3791221
ETNPPL, rs7687209
KNG1, rs11919484
SNAPC4, rs10122824
RASD1, rs1242484
METTL7A, rs7966357
LEAP2, rs12163971
METTL10, rs10430743
KLHDC7A, rs11261022
TRIM6, rs396341
AMIGO1, rs407102
FGF5, rs1458038
GALNTL5, rs10224002
UMOD, rs77924615
Locus
Renal Tubulointerstitial
Renal Glomerular
Adipose Subcutaneous
Adipose Visceral Omentum
Adrenal Gland
Artery Aorta
Artery Coronary
Artery Tibial
Brain Caudate Basal Ganglia
Brain Frontal Cortex Ba9
Brain Hypothalamus
Brain Nucleus Accumbens Basal Ganglia
Breast Mammary Tissue
Cells Ebv Transformed Lymphocytes
Cells Transformed Fibroblasts
Colon Sigmoid
Colon Transverse
Esophagus Mucosa
Esophagus Muscularis
Heart Atrial Appendage
Heart Left Ventricle
Liver
Lung
Muscle Skeletal
Nerve Tibial
Ovary
Pancreas
Pituitary
Prostate
Skin Not Sun Exposed Suprapubic
Skin Sun Exposed Lower Leg
Small Intestine Terminal Ileum
Spleen
Stomach
Testis
Thyroid
Uterus
Vagina
Whole Blood
Posterior probability of colocalization
[0.0, 0.2)
[0.2, 0.4)
[0.4, 0.6)
[0.6, 0.8)
[0.8, 1.0]
Change in gene expression with decreased eGFR
0.029
0.010
0.005
0.000
−0.005
−0.010

Figure 6 – Co-localization of independent eGFR-association signals at the *UMOD*/*PDILT* locus with urinary uromodulin concentrations supports *UMOD* as the effector gene.

Association plots: association $-\log_{10}$(p-value) (Y axis) vs. chromosomal position (X axis). Approximate conditional analyses among EA individuals support the presence of two independent eGFR-associated signals (**Panel A**). The association signal with urinary uromodulin/creatinine levels looks similar (**Panel B**). Co-localization of association with eGFR (upper sub-panel) and urinary uromodulin/creatinine levels (lower sub-panel) for the independent regions centered on *UMOD* (**Panel C**) and *PDILT* (**Panel D**) support a shared underlying variant in both regions with high posterior probability.


A association with eGFR
rs77924615
rs34882080
−log10(p−value)
Recombination rate (cM/Mb)
GP2
UMOD
PDILT
ACSM5
ACSM2A
Position on chr16 (Mb)
B association with uromodulin
rs34262842
rs77924615
C region 1: UMOD
eGFR
rs34882080
rs13334589
H4 pp=0.97
uromodulin
conditional −log10(p−value)
D region 2: PDILT
rs77924615
H4 pp=0.96